

WHAT WE TEST WHEN WE TEST FOREIGN LANGUAGE READING SKILLS

1. Introduction

Until the early 1980s the major concern of testers, teachers and researchers was with the issue of methodology, i.e. how we test reading skills. In the 1980s one could note a growing interest in the nature of the reading construct itself, i.e. what we test when we test reading skills. In other words, a move away from a focus on method to a focus on the content of reading tests. Nowadays it is assumed that once we are aware of “the performance conditions that need to be built into a test and the reading skills and strategies that are to be tested”, only then can we make decisions on appropriate testing techniques (Urquhart and Weir 1998: 121).

The article discusses the nature of the reading construct, i.e. what we test when we test reading. It explains the difference between reading interpretation and reading comprehension in relation to testing. The assumption is put forward that reading tests cannot account for individual interpretations of texts; they can measure reading comprehension, i.e. understanding directed by test tasks. The article also presents several taxonomies of reading skills and discusses their usefulness in writing syllabuses and constructing tests. Special attention is devoted to Weir’s (1993) taxonomy, who breaks down reading into four reading types. The article discusses the status of each reading type in tests applied nowadays and examines how these types are operationalised. Finally, the article offers some advice how to construct good reading comprehension tests.

2. The limits of testing reading skills

At the beginning of the discussion about testing reading skills, it is crucial to consider the question: why we read. There are various reasons why we read: we may read to gain a general background in an area, for amusement, aesthetic satisfaction, or simply to browse. Would it be possible to measure enjoyment derived from a novel? We could measure how much someone remembers from the plot, details, etc., but not someone’s aesthetic satisfaction derived from reading. It is worth realising that the

more we are aware of the nature of reading that we attempt to test, the more we are likely to understand the limits of testing reading.

At this moment it is also useful to make a distinction between interpretations and comprehension (Urquhart 1987). It seems that in reading tests we test comprehension, not interpretations. The same text may be read in many different ways. The process of reading is an interaction between the reader and the text so there may be different readings of the same text, i.e. different **interpretations**. Readers from different cultures (either ethnic or professional) will understand the same text differently. The same reader at different times with different knowledge will read the same text in a different way.

Candlin (1984) concludes that conventional tests cannot make allowance for individual interpretations. Tests should test **comprehension**, i.e. a state of achievement, a product of understanding which is controlled by specific test tasks, requiring specific ways of reading. Different test questions aim to elicit different reading behaviours; e.g., some intend to make readers read for the gist, others – read carefully. And it is such reading behaviours elicited by test questions that tests can measure.

3. Reading taxonomies

Reading taxonomies are lists of skills that constitute the reading process. Reading skills can be described as cognitive abilities which the reader is able to apply when interacting with written texts. By some researchers, e.g. Lunzer et al. (1979), skills are recommended as the best framework for structuring reading syllabuses, writing test specifications and constructing tests. Below two taxonomies are presented and their usefulness in designing tests is discussed.

The taxonomy by Davis (1968), as cited by Urquhart and Weir (1998: 90), distinguishes the following skills:

- * Identifying word meanings.
- * Drawing inferences.
- * Identifying writer's technique and recognising the mood of the passage.
- * Finding answers to questions.

The taxonomy by Munby (1978), as cited by Urquhart and Weir (1998: 90), presents the following reading components:

- * Recognising the script of a language.
- * Deducing the meaning and use of unfamiliar lexical items.
- * Understanding explicitly stated information.
- * Understanding information when not explicitly stated.
- * Understanding conceptual meaning.
- * Understanding the communicative value of sentences.
- * Understanding the relations within the sentence.
- * Understanding relations between parts of text through lexical cohesion devices.
- * Interpreting text by going outside it.
- * Recognising indicators in discourse.
- * Identifying the main point of information in discourse.

- * Distinguishing the main idea from supporting detail.
- * Extracting salient points to summarise the text, an idea.
- * Selective extraction of relevant points from a text.
- * Basic reference skills.
- * Skimming.
- * Scanning to locate specifically required information.
- * Transcoding information to diagrammatic display.

It is easy to criticise Davis' taxonomy. The skill of "Finding answers to questions" seems to include all the others. Munby's taxonomy appears very comprehensive. In fact it has been very popular; many test specifications have been written on the basis of this taxonomy. For example, the testers from the Teacher Training College of the Jagiellonian University used this taxonomy while developing the reading subtest of the standardised Final Practical Exam in the Kraków cluster of teacher training colleges (see Kusiak and Jurek-Kwiatkowska 1998).

A very practical question arises here: whether the reading process can be broken down into separate skills; and consequently whether we can be sure that particular test items test only and nothing more than we think they test. Most reading items test a variety of skills and different readers bring different processes to bear on a text. It is relevant to mention a study of Alderson (1990), who found that when students introspected on their test-taking techniques they sometimes seemed to use different processes from each other and yet reached the same answer. In fact researchers give evidence against both fully unitary views of reading ability on the one hand and multidimensional views on the other. However, it seems that for the purposes of teaching and testing a view that accepts breaking reading into separate skills is more useful.

This view is seen in the recent taxonomy suggested by Urquhart and Weir (1998: 123), based on Weir (1993) and Pugh (1978). Four reading types are distinguished in this taxonomy: global expeditious, global careful, local expeditious and local careful reading (see the table below).

Table 1

Reading types as suggested by Urquhart and Weir (1998: 123)

	Global	Local
Expeditious	Skimming quickly to establish discourse topic and main ideas. Search reading to locate quickly and understand information relevant to pre-determined needs.	Scanning to locate specific information; symbol or group of symbols; names, dates, figures or words.
Careful	Reading carefully to establish accurate comprehension of the explicitly stated main ideas the author wishes to convey; propositional inferencing.	Understanding syntactic structure of sentence and clause. Understanding lexical and/or grammatical cohesion. Understanding lexis/ deducing meaning of lexical items from morphology and context.

Global comprehension refers to comprehension beyond the level of micropropositions – from macropropositions to discourse topic. Local comprehension refers to the decoding of micropropositions and the relations between them. In careful reading and in skimming the reader makes a conscious effort to construct a macrostructure, the gist of the text. In careful reading, this is likely to be done by reference to the whole text, in skimming from parts of the text. In scanning there is no attempt to construct a macrostructure, in search reading it is probable that only certain key ideas in the macrostructure will be sought. It is worth noting that whether we measure global or expeditious reading depends on the time given for a test. Expeditious reading is quick and selective reading. Tests measuring this type of reading should take less time than tests measuring careful reading.

When we compare this matrix with the other taxonomies discussed above, it seems that the distinction into four reading types is less detailed, clearer and thus more helpful in test development. The Weir taxonomy makes it easier to analyse the reading construct of reading test types.

Skimming tests (asking questions, e.g. “What is the text about?”, “Suggest the title for the given passage”) test an ability to establish a superordinate macroproposition, a discourse topic for a text, thus global expeditious reading. Items measuring understanding the main idea (asking questions: “What is the main idea of the paragraph?”) test global careful and global expeditious reading. Scanning items, questions checking reading for specific information, gap filling items and cloze tests measure expeditious and careful reading skills on the micropropositional local level.

The matrix of four reading types encourages testers to analyse the structure of texts. It enables them to distinguish macropropositions and micropopositions of the text. An understanding of the relationship between micro- and macropropositions in text is important for the test developer in order to establish the level of focus of test items. Equipped with a structured description of the content of a text, test developers are in a better position to construct valid items.

The matrix discussed above enables testers to consider a wider range of reading behaviours. It can help test writers to decide what reading types they want to test. It is particularly useful in constructing exams that consist of a number of tests aiming to measure a range of reading skills.

The taxonomy of Weir pays equal attention to careful and expeditious reading. The overview of exam tests and tests used in reading studies over the last 50 years done by Urquhart and Weir (1998) points to the fact that both exam tests and tests applied as research instruments failed to include items testing expeditious reading. They were mainly concerned with careful reading, and with extracting complete meanings from the text as opposed to comprehension that aims at only extracting main ideas, skimming or searching for particular details. In this situation the elucidation of the role of the expeditious type of reading in test construction seems particularly important. This issue is developed below.

4. The status of expeditious reading in teaching and testing reading

Weir (1983) claims that expeditious reading, i.e. reading quickly and efficiently in the target language, can serve as a better discriminator between L1 and L2 readers. Urquhart and Weir (1998: 130) state that: "Slow careful reading also poses problems but the difference between L1 readers and L2 readers is most marked in expeditious reading". However, as has been pointed out before, expeditious reading is neglected in testing and research.

The low status of expeditious reading may be caused by the following factors. It may be the belief that there is something like "perfect" comprehension of a text and the assumption that careful reading, which aims to extract perfect comprehension, is superior to any other kind of reading, e.g. skimming. Another assumption may be the notion that the "ideal" comprehension consists of the recovery of "author's meaning". And only careful study of text allows full understanding of author's meaning. Such beliefs lead to underestimation of other reading types, e.g. scanning, in which the reader may pay little attention to the author's intentions. Changing these views on reading brings serious consequences both for the teaching and testing of reading.

My observations (Kusiak 1999) of how English is taught in Polish schools show that teaching and testing contexts strengthen the status of careful reading, i.e. study of texts. In teaching short texts are favoured, and as a result only intensive classroom reading is practised, and extensive independent reading done out of class is neglected. It may be often observed that teachers prefer to use texts for language study rather than for developing reading skills. Besides, students are often asked to read texts without being given any task/question that would specify the purpose of their reading. This teaching procedure results in reading the text in a careful way. It is worth stressing that this is often done against the philosophy of contemporary coursebooks, which offer different reading tasks to teach different readings.

Testing reading is not a popular procedure in Polish schools. Since it is a part of the final "Matura" exam or the Cambridge exams, it is practised usually when students revise for the exams. Tests do not promote real life reading situations. They contain texts consisting of several short paragraphs; often they are disembodied prose without titles, pictures or headings typical of texts encountered in real life. They are typed, which deprives them of original text layout. Because of practical reasons, i.e. problems with administration of tests, they often do not test expeditious reading.

5. Factors to consider in testing reading

Presented below are suggestions that I find useful in constructing reading tests. The ideas are based on the advice given by Weir (1993), and Urquhart and Weir (1998). In this discussion the taxonomy of reading types by Weir (1993) is referred to.

It is very important to consider the amount of time students are given in order to complete the test. The amount of time students spend on the test items may change the operations that are needed to answer them. Testing expeditious reading requires less

time than testing careful reading. This would mean timing students and collecting papers after, e.g. five minutes. Too much time given for testing expeditious reading may change the test into a test of careful reading; too little time given for a careful test will change it into a test of expeditious reading. If it is not feasible to collect expeditious reading tests once they have been completed, testers can place this section as the last one in the test paper. It is a procedure adopted by, e.g. FC exam tests.

A good idea would be not to combine items/sections testing careful reading and items/sections testing expeditious reading in the same test paper based on the same text. This will strengthen the validity of test items included in the test.

Much depends on the length of texts used in the tests. Too short texts may be artificial and difficult for testing expeditious reading, although they would be suitable for testing careful reading. I have noticed that both teachers and testers are reluctant to use long texts. This may be due to practical factors, such as photocopying expenses as well as the belief that using long texts “only” to practise expeditious reading may be a waste of the resources. I would like to encourage testers to use long texts; such texts resemble real-life reading situations and whether used for testing careful or expeditious reading they make reading tests very authentic.

We should also consider the difficulty of texts. More difficult texts with unclear dense organisation or difficult vocabulary may be good for testing general reading and testing main ideas.

When designing tests, it is of vital importance to choose appropriate testing techniques. Below the most typical test types are discussed.

Multiple choice questions are good at testing isolated details or “fragmentary” comprehension. They are not very suitable for testing global reading, such as a broader response to the text, comprehension across the text, understanding the writer’s intention, the overall message or the structure of the text. It is true that skimming and predicting may be tested by means of multiple choice questions, but we are never sure whether we really test these skills because we provide options for readers; readers keep them in mind while processing the text.

For global comprehension tests simple open-ended questions may be better. We are more sure that we test what we want to test. Short answer questions lend themselves to testing all types of reading: search reading, skimming for gist, scanning for specific information and reading carefully to extract the main ideas and important details. Recent testing research, e.g. Carrell, Pharis and Liberto (1989), recommends short answer questions because of their good discriminating qualities.

Information transfer tests are suitable for testing all types of reading; e.g. extracting specific information and careful reading of more complex skills.

Summary tests require completing a passage that is a summary of the text. There is a danger that completion may involve only transferring words from the text. It is doubtful that a full range of reading operations would be triggered. It would be difficult to test skimming by means of such tests.

Identifying appropriate headings for parts of the text is good to test skimming. Gapped texts requiring inserting missing paragraphs or sentences are suitable for testing cohesion, i.e. understanding relations between different parts of text.

6. Conclusion

The purpose of this paper was to highlight the need to consider the construct of reading tests before choosing texts and deciding on appropriate testing techniques. By thinking a priori about reading skills that are to be tested test developers can construct more valid tests. This awareness could also help teachers. It can make them more selective about tests offered in the coursebooks. Last, but not least, it could make teachers be more critical about their own teaching.

BIBLIOGRAPHY

- Alderson, C. 1990. "Testing reading comprehension skills". *Reading in a Foreign Language* 7: 465–503.
- Candlin, C. 1984. *Preface*. In: Alderson, C. and Urquhart, A. (eds.) *Reading in a Foreign Language*. London: Longman, ix–xiii.
- Carrell, P., Pharis, B. and Liberto, J. 1989. "Metacognitive strategy training for ESL reading". *TESOL Quarterly* 23: 647–678.
- Davis, F. 1968. "Research in comprehension in reading". *Reading Research Quarterly* 3: 499–545.
- Kusiak, M. 1999. "Interakcyjne techniki nauczania sprawności czytania w języku obcym". *Neofilolog* 18: 40–47.
- Kusiak, M. and Jurek-Kwiatkowska, L. 1998. "Towards standardisation of the final practical English examination in the Kraków cluster of colleges". In: Melia, P. (ed.) *Innovations and Outcomes in English Language Teacher Education*. Warszawa: The British Council Poland, 231–239.
- Lunzer, E., Waite, M. and Dolan, T. 1979. "Comprehension and comprehension tests". In: Lunzer, E. and Gardner, K. (eds.) *The Effective Use of Reading*. London: Heinemann Educational, 37–71.
- Munby, J. 1978. *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Pugh, A. 1978. *Silent Reading*. London: Heinemann Educational.
- Urquhart, A. 1987. "Comprehensions and interpretations". *Reading in a Foreign Language* 3: 387–409.
- Urquhart, A. and Weir, C. 1998. *Reading in a Second Language: Process, Product and Practice*. London: Longman.
- Weir, C. 1983. *Identifying the Language Needs of Overseas Students in Tertiary Education in the United Kingdom*. Ph.D. thesis. University of London, Institute of Education.
- Weir, C. 1993. *Understanding and Developing Language Tests*. London: Prentice Hall.